

DOCUMENT RESUME

ED 109 180

TM 004 628

AUTHOR
TITLEHutchinson, Thomas E.
Methodological Developments to Monitor Process of
Some Sources of Error in the Efficiency of
Evaluations.PUB DATE
NOTE[Apr 75]
10p.; Paper presented at the Annual Meeting of the
American Educational Research Association
(Washington, D. C., March 30-April 3, 1975)EDRS PRICE
DESCRIPTORSMF-\$0.76 HC-\$1.58 PLUS POSTAGE
*Decision Making; *Efficiency; *Error Patterns;
*Evaluation Methods; *Evaluation Needs

ABSTRACT

Specific evaluation design is a non trivial problem which must be developed through a set of procedures involving considerable study of the enterprise to be evaluated. Three criteria to be considered when an evaluator provides data for a decision maker are: (1) the percentage of the evaluation data that are actually used by the decision maker; (2) the percentage of the decision maker's decisions that are made, in part, with the use of the evaluation data; and (3) the degree to which the data is utilized in the most important decisions, even where all decisions are not made using the evaluation data. These criteria have been named efficiency, completeness, and focus, respectively. Evaluators should avoid errors, and prefer operational to non operational techniques. Evaluator/decision maker interaction patterns should be incorporated into the evaluation design. (Author/BJG)

* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *

ED109180

Methodological Developments to Monitor Process

or

Some Sources of Error in the Efficiency of Evaluations

by Thomas E. Hutchinson
University of Massachusetts

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY.

TM 004 628

Paper delivered at the AERA Convention,
Washington, D.C., April, 1975

Introduction

In recent years there has been considerable activity in the development of methodology for evaluation (Scriven (1967), Stufflebeam et al. (1971), Provus (1971), Benedict (1973a)). Most of these new systems propose a process for the development of an evaluation design which is specific to the project, program, curriculum, or enterprise to be evaluated. The process of designing the evaluation is viewed as part of the evaluation activity rather than something that precedes the beginning of evaluation.

Although there are important differences among these new approaches, most are in agreement that the specific evaluation design is a non trivial problem which must be developed through a set of procedures involving considerable study of the enterprise to be evaluated. Evaluators using such approaches find it difficult to respond to requests for proposals which require the proposal to specify the specific evaluation design and, sometimes, even the measuring instruments to be used. In order to respond they either have to do one fourth or more of the total work for nothing and with no guarantee of being accepted for the remainder of the work or abandon their professional beliefs and propose some pre-designed plan like a pre-post experimental group--quasi control group design, using some specified standardized instruments as the measures to be employed. Many of these evaluators see the use of pre-designed techniques, either in the design or in instrumentation, as inappropriate to the purpose of evaluation.

Inappropriateness of Pre-Designed Techniques

A large number of evaluators have as the purpose for their work: to provide data for decision making (Stufflebeam (1969)). Some of these

evaluators define this purpose in terms of three criteria: the percentage of the evaluation data that are actually used by the decision maker to whom they are provided; the percentage of a decision maker's decisions that are made, in part, with the use of evaluation data; and where not all decisions are made using evaluation data the degree to which the data are provided for the more important decisions (Hutchinson (1972)).

These criteria have been named, efficiency, completeness, and focus respectively. Coffing, et al. (1973) have provided operational procedures for measuring these criteria. Completeness is the percentage of the decisions of a decision maker that are made, in part, with the use of evaluation data. Each decision maker is asked to keep a log recording each decision that he or she makes and what data, if any, were used in making the decision. The evaluator keeps a log of the evaluation data that are provided to each decision maker. At regular intervals the evaluator by inspection of the log of decisions can determine whether or not each decision was made, in part, with the use of evaluation data. The percent of the decisions that were made using evaluation data is the percentage of completeness.

The percentage of completeness that can be obtained is limited by the amount of resources available to provide the decision maker with data. These evaluators are trying to maximize the percentage of completeness within the resource constraints. Less than this represents error of completeness.

Where the percentage of completeness is less than 100%, focus is the degree to which data were provided for the more important decisions. At regular intervals the evaluator can ask the decision maker to prioritize the decisions on the log by the criteria of importance. The evaluator

can construct a two by two contingency table with categories of evaluation data used, evaluation data not used and most important, least important. The marginal for used, not used is the marginal imposed for most important, least important. A percentage of correct focus may then be calculated. These evaluators are trying to design an evaluation that will be 100% focused on the most important decisions. Less than this represents error of focus.

Efficiency is the percentage of the evaluation data which are used in making decisions. At regular intervals the evaluator by inspection of the log of decisions can determine whether or not each evaluation data report, variable by variable, has been used by the decision maker in making any decisions. The percent of the data provided that were actually used is the percentage of efficiency. These evaluators are trying to design an evaluation that will have 100% efficiency. Less than this represents error of efficiency.

The use of pre-designed techniques is potentially a considerable source of error of efficiency. In the case of the quasi control group design the decision maker may reject the usefulness of the data on the basis that the control group is either known to be (or simply might be) different in some systematic way from the experimental group. In the case of the use of a pre-designed measuring instrument; for example, a standardized self concept test; the decision maker may find the data not to be useful because upon inspection of the test the decision maker finds that it does not have face validity with respect to his or her meaning of self concept. It is only when the specific evaluation design is worked out with the decision maker that the evaluator can be sure that these threats to the efficiency of the evaluation design can be avoided.

The potential sources of error of efficiency just mentioned can be categorized as bias or as a lack of decision maker validity. Further, in these situations the decision maker was aware of the lack of decision maker validity. It is possible under some circumstances, for there to be a lack of decision maker validity and for the decision maker to be unaware of this bias. In this event the data will be used when they shouldn't be used.

Non Operational Techniques

This kind of error is possible when the measuring instrument is not operational. Imagine that an evaluator is attempting to provide a project director with data for decision making. The project director has agreed to having the evaluator do direct observation in the classrooms using a rating form. One item on the rating form is as follows:

The students are working in teams -- none, some,
about half, most, all.

If the project director and the evaluator observed the same classroom at the same time they might not agree on the observation. The evaluator might count the number of students working alone. The project director might count the number of students in groups of three or more. The two observers have different ideas about what constitutes a team. The evaluator would report more team activity than is accurate from the perspective of the decision maker. The project director might decide that everything is as he wants it when everything is not as he wants it.

An operational item would not have this problem. Consider the following items:

Count the number of students in groups of three or more.

Count the number of students in the room.

These items are not subject to misinterpretation or misunderstanding.

The unoperational item is also subject to certain other types of error of measurement. Consider again the item:

The students are working in teams -- none, some,
about half, most, all.

One observer may believe that school work is quiet and not consider noisy groups to be "working in teams." A different observer may believe that if it is quiet then the students are not really "working in teams" but rather individually in a group. Therefore, this observer would not count quiet groups. Non operational items permit observer reactivity.

Another possibility is that the observer begins with the first classroom assuming that noisy groups aren't really working. By the end of the day the noise has really gotten to the observer and now only totally silent groups get counted. Due to this observer maturation not all classrooms have had the same standard applied.

Consider a second item on the same rating form. It is as follows:

The students are cooperating with each other --
none, some, about half, most, all.

The Observer who believes that school work should be quiet also believes that the quiet students are cooperating with each other. Therefore, this observer counts the quiet groups and quiet isolates. Since both items are judged by a "level of noise" standard the results are that the two measures are not totally independent. The other observer who likes vigorous student interaction scores both items the same because cooperating is interaction in a group. For this observer the items are totally dependent. The project director is getting one piece of information although he thinks he is getting two separate pieces of information.

The use of operational items avoids these kinds of problems. The evaluator has the decision maker operationally define the intents or goals of the program. Techniques can then be developed to measure the operational items.

Procedures for Avoiding Errors

Procedures have been developed for providing data for decision making in such a way that the above kinds of errors are avoided. The Fortune/Hutchinson Evaluation Methodology (Benedict (1973), (1974)) provides for regular interaction between the decision maker and the evaluator in the design of the evaluation. The procedures require the approval of the decision maker at the end of each phase of design.

This methodology also provides for operationally defining the decision maker's intents (Hutchinson and Benedict (1970)) so that operational measurements can be designed. These procedures are designed in an attempt to maximize the criteria for efficiency, completeness, and focus. They also provide a process for evaluating the evaluation design itself in terms of these criteria.

The F/H Evaluation Methodology has been subjected to methodological research by Jones (1971), Gordon (1973), Benedict (1973b), Rosen (1974), and Jeffers (1974). In each case the intent of the research was to identify problems with the procedures. In each case problems were found and solutions have been proposed or are under development. Unfortunately, a proposed symposium to present the results of this methodological research on evaluation was turned down by Division D of AERA.

17-

References Cited

- Benedict, Larry G. The Fortune/Hutchinson evaluation methodology: a decision oriented approach. A symposium paper presented at AERA, New Orleans, 1973a.*
- Benedict, Larry G. The goals gap in educational evaluation: identification and development of methodology. Unpublished doctoral dissertation, University of Massachusetts, 1973b.
- Benedict, Larry G. Practical Guide for Evaluation, Capital Region Educational Council, Hartford, Conn. 1974*
- Coffing, R. T., Hodson, W. A., and Hutchinson, T. E. A needs analysis methodology for education of the handicapped -- Version I. Final Report of Connecticut Project No. BB-404. North Haven, Connecticut: Area Cooperative Educational Services, 1973.*
- Gordon, G. M. A field test of the Fortune/Hutchinson Evaluation methodology as it could be employed in the evaluation of national urban league street academies. Unpublished doctoral dissertation, University of Massachusetts, 1973.
- Hutchinson, T. E., and Benedict, L. G. The operationalization of fuzzy concepts. University of Massachusetts, mimeo, 1970.*
- Hutchinson, T. E. Some overlooked implications of the purpose: to provide data for decision making. A symposium paper presented at AERA, Chicago, 1972.*
- Jeffers, Ellaine. The interface of measurement and evaluation: methodological development and field test in an innovative environment. Unpublished doctoral dissertation, University of Massachusetts, 1974.
- Jones, L. The operationalization of Educational objectives for the evaluation of an on-going program. Unpublished doctoral dissertation, University of Massachusetts, 1971. *
- Provus, M. Discrepancy evaluation for educational program improvement and assessment. Berkeley, California: McCutchan, 1971.
- Rosen, David. The Shanti evaluation: a study of the Fortune/Hutchinson evaluation methodology in a public alternative school Unpublished doctoral dissertation, University of Massachusetts, 1974.

Scriven, M. The methodology of evaluation. In R. W. Tyler (Ed.), Perspectives of curriculum evaluation, AERA Curriculum Evaluation Monograph Series, #1. Chicago: Rand McNally, 1967.

Stufflebeam, D. L. Evaluation as enlightenment for decision making. In W. Beatty (Ed.), Improving educational assessment. Washington, D. C.: Association for Supervision and Curriculum Development, 1969.

Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., and Provus, M. M. Educational evaluation and decision making, Itasca, Illinois: F. E. Peacock Publishers, 1971.

* Copies of these papers may be acquired by writing to the author of this paper.